

Linking routinely collected social work, education and health data to enable monitoring of the health and health care of school-aged children in state care ('looked after children') in Scotland: a national demonstration project

Clark, D.; King, A.; H Sharpe, K.; Connelly, G.; Elliott, L.; Macpherson, L.M.D.; McMahon, A.D.; Milligan, I.; Wilson, P.; Conway, D.I.; Wood, R.

Published in:
Public Health

DOI:
[10.1016/j.puhe.2017.05.003](https://doi.org/10.1016/j.puhe.2017.05.003)

Publication date:
2017

Document Version
Author accepted manuscript

[Link to publication in ResearchOnline](#)

Citation for published version (Harvard):

Clark, D, King, A, H Sharpe, K, Connelly, G, Elliott, L, Macpherson, LMD, McMahon, AD, Milligan, I, Wilson, P, Conway, DI & Wood, R 2017, 'Linking routinely collected social work, education and health data to enable monitoring of the health and health care of school-aged children in state care ('looked after children') in Scotland: a national demonstration project', *Public Health*, vol. 150, pp. 101–111.
<https://doi.org/10.1016/j.puhe.2017.05.003>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please view our takedown policy at <https://edshare.gcu.ac.uk/id/eprint/5179> for details of how to contact us.

Title

Linking routinely collected social work, education, and health data to enable monitoring of the health and healthcare of school aged children in state care ('looked after children') in Scotland: a national demonstration project.

Authors

David Clark¹, dclark5@nhs.net

Albert King², albert.king@gov.scot

Katharine Sharpe³, kathy.h.sharpe@btinternet.com

Graham Connelly⁴, g.connelly@strath.ac.uk

Lawrie Elliott⁵, lawrie.elliott@gcu.ac.uk

Lorna MD Macpherson³, lorna.macpherson@glasgow.ac.uk

Alex D McMahon³, alex.mcmahon@glasgow.ac.uk

Ian Milligan⁴, ian.milligan@strath.ac.uk

Philip Wilson⁶, p.wilson@abdn.ac.uk

David I Conway^{1,3}, david.conway@glasgow.ac.uk

Rachael Wood^{1,7}, rachaelwood@nhs.net

1 Information Services Division, NHS National Services Scotland, Gyle Square, 1 South Gyle Crescent, Edinburgh, EH12 9EB

2 Education Analytical Services Division, Scottish Government, Victoria Quay, Edinburgh, EH6 6QQ

3 School of Medicine, Dentistry, and Nursing, University of Glasgow, 378 Sauchiehall Street, Glasgow, G2 3JZ

24 4 CELSIS (Centre for Excellence for Looked After Children in Scotland), University of
25 Strathclyde, Curran Building, 94 Cathedral Street, Glasgow, G4 0LT
26 5 Department of Nursing and Community Health, School of Health and Life Sciences
27 Glasgow Caledonian University, Cowcaddens Road Glasgow, G4 OBA
28 6 Centre for Rural Health, University of Aberdeen, Old Perth Road, Inverness, IV2 3JH
29 7 Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh,
30 EH8 9AG
31
32 Corresponding author:
33 Rachael Wood
34 Information Services Division
35 NHS National Services Scotland
36 Gyle Square
37 1 South Gyle Crescent
38 Edinburgh
39 EH12 9EB
40 rachaelwood@nhs.net
41 0131 275 7028

Abstract

Background and objectives

Children in state care ('looked after children') have poorer health than children who are not looked after. Recent developments in Scotland and elsewhere have aimed to improve services and outcomes for looked after children. Routine monitoring of the health outcomes of looked after children compared to those of their non-looked after peers is currently lacking. Developing capacity for comparative monitoring of population based outcomes based on linkage of routinely collected administrative data has been identified as a priority. To our knowledge there are no existing population based data linkage studies providing data on the health of looked after and non-looked after children at national level. Smaller scale studies that are available generally provide very limited information on linkage methods and hence do not allow scrutiny of bias that may be introduced through the linkage process.

Study design and methods

National demonstration project testing the feasibility of linking routinely collected looked after children, education, and health data.

Participants

All children in publicly funded school in Scotland in 2011/12.

Results

Linkage between looked after children data and the national pupil census classified 10,009 (1.5%) and 1,757 (0.3%) of 670,952 children as, respectively, currently and previously looked after. Recording of the unique pupil identifier (Scottish Candidate Number, SCN) on looked after children returns is incomplete, with 66% of looked after records for 2011/12 for children of possible school age containing a valid SCN. This will have resulted in some under-ascertainment of currently and, particularly, previously looked after children within the

general pupil population. Further linkage of the pupil census to the NHS Scotland master patient index demonstrated that a safe link to the child's unique health service (Community Health Index, CHI) number could be obtained for a very high proportion of children in each group (94%, 95%, and 95% of children classified as currently, previously, and non-looked after respectively). In general linkage rates were higher for older children and those living in more affluent areas. Within the looked after group, linkage rates were highest for children with the fewest placements and for those in permanent fostering.

Conclusions

This novel data linkage demonstrates the feasibility of monitoring population based health outcomes of school aged looked after and non-looked after children using linked routine administrative data. Improved recording of the unique pupil identifier number on looked after data returns would be beneficial. Extending the range of personal identifiers on looked after children returns would enable linkage to health data for looked after children who are not in publicly funded schooling (i.e. those who are pre- or post-school, home schooled, or in independent schooling).

Word count (431)

Keywords

Child in care

Looked after children

Delivery of healthcare

Dental health services

Medical record linkage

Public health informatics

Scotland

Introduction

In Scotland, children in state care (referred to as ‘looked after children’) are those under supervision or accommodated by local authorities¹⁻³. Children can become looked after following a voluntary agreement with their parents or a compulsory process involving the Scottish Children’s Hearing System⁴ or the courts, and their requirement for such support can reflect care, protection, and/or offending needs. Looked after children may live at home with their parents under social work supervision (‘looked after at home’); with other family members or friends (‘kinship care’); with foster carers or prospective adopters; or in residential accommodation provided by the state (residential units, schools, and secure care)⁵.

Around 15,400 children in Scotland were looked after at the end of July 2015, around 1.5% of all children aged less than 18 years⁶. The health, educational, and wider social outcomes of looked after children are generally poorer than those of children who are not looked after⁷⁻⁹. Current Scottish Government policy strongly supports improving the experience and outcomes of looked after children¹⁰ and emphasises the need for robust routine data to enable monitoring of care provided and outcomes achieved¹¹.

Currently, routine data returned by local authorities to the Scottish Government on children being looked after form the basis of an annual statistical publication on children’s social work⁶. In addition, the Scottish Government routinely links the looked after data to administrative data returned by local authorities on education provision to enable monitoring of the educational attainment and post school destinations of looked after children compared to all children¹². Scotland has a wide range of high quality routine health data that can be used to monitor child health. Health records in general do not include information on

children's looked after status hence they cannot be used in isolation to assess the health of looked after children. Linkage of routine looked after and health data would open up the possibility of robust population based monitoring of the health outcomes of looked after and non-looked after children, and developing such a linkage has been identified as a priority¹¹.

Here we report the results of a national level demonstration project linking routinely available looked after children data and health data for the first time in Scotland. To our knowledge, this is the first time globally that such a national level, population based linkage study has been undertaken. This paper reports the methodology and results of the linkage process: a separate paper reports the results of a follow on analysis assessing the dental health of looked after and non-looked after children using the linked data set created (submitted for publication, available on request). With this paper we aim to provide information of use to future researchers wishing to assess the health and healthcare of looked after children using routinely available data.

Methods

We made use of three databases. These are described in full in Panel 1 and were (a) annual looked after data returns submitted by local authorities to the Scottish Government providing information on all children looked after by the local authority during the preceding academic year ('looked after data'), (b) the annual pupil census (again returned by local authorities to the Scottish Government) which is the master index of all children in publicly funded schools in Scotland at the start of each academic year ('pupil census'), and (c) the Community Health Index database which is the master patient index for the NHS in Scotland, continuously maintained primarily from General Practitioner registration data ('CHI database'/'health data'). Publicly funded schools comprise mainstream and special schools funded by the state

and managed by local authorities. Children who are home schooled or attending an independent or charitable day or residential school or secure unit will not be included in the pupil census, even if their school place is being funded by their local authority. The pupil census contains the Scottish Candidate Number (SCN) for every pupil. The SCN is the unique identifier used on all education records in Scotland. The looked after data should contain the SCN if it is available, for example if the child is attending publicly funded school. The Community Health Index database contains patients' CHI number, the unique identifier used on all health records in Scotland.

Panel 1 about here.

Linkage of records belonging to the same child across databases depends on the personal identifiers held within the databases. There are insufficient personal identifiers (date of birth and gender only) in the looked after data to enable robust direct linkage to the CHI database. We therefore used Scottish Candidate Numbers (SCNs) included in looked after data from 2007/08 to 2011/12 to identify looked after children in the 2011/12 pupil census. We then used the wider range of personal identifiers available within the pupil census (date of birth, gender, and home postcode) to link this database to the CHI database. This resulted in a look up file providing an SCN (with a flag to indicate looked after status) to CHI number look up key for all children in publicly funded school in 2011/12 (see Figure 1). Further details of the linkage process are provided as supplementary material. Note that the limited personal identifiers available in looked after data returns mean that linking to the CHI database must go through this intermediate step of the pupil census. It is currently not possible to link looked after records for pre- or post-school children to their health records.

Figure 1 about here.

Categorisation of looked after status in the pupil census relies on accurate recording of children's SCNs within the looked after data. We therefore first assessed the quality of recording of SCN on looked after data by examining the proportion of records for children of school age that contained any SCN or a valid SCN. The looked after data validation rules define a valid SCN as one of the correct length (nine digits) that contains a correct check digit.

We then assessed the proportion of currently, previously, and non-looked after children in the 2011/12 pupil census whose record could be safely linked to the child's CHI number. Using demographic information contained within the pupil census we further assessed linkage rates by pupil age group, gender, Scottish Index of Multiple Deprivation quintile, ethnicity, and by school local authority area. The Scottish Index of Multiple Deprivation is a small area based measure of deprivation: individuals are assigned to deprivation quintiles based their postcode of residence¹³. Finally, using information contained within looked after data we assessed linkage rates for currently looked after children by the cumulative number of looked after placements they had experienced (placements ending on or after 1st August 2007) and the type and legal reason underpinning their most recent placement. Linkage rates for currently and, separately, previously looked after children were compared to those for non-looked after children using 95% confidence intervals for the difference in two independent proportions. Similarly, linkage rates for particular categories of currently looked after children were compared to those in a relevant reference category (one placement since August 2007; looked after at home; and compulsory supervision through the Children's Hearing system (resident at home)).

Once the linkage was completed, CHI numbers were used to identify and extract a range of routinely available dental health and healthcare data held by the NHS Scotland Information Services Division (ISD) to enable comparison of the dental health of school aged looked after children to that of their non-looked after peers. The results of this analysis are reported separately (submitted for publication, available on request). Approval for this study was obtained from the Scottish Government Education Analytical Services Division Data Access Panel and the NHS Privacy Advisory Committee. Ethical approval was obtained from the University of Glasgow College of Medicine, Veterinary, and Life Sciences Ethics Committee: NHS ethical approval was not required.

Results

The quality of recording of Scottish Candidate Number (SCN) on looked after records for school aged children is shown in Table 1. A total of 20,771 children were included in the 2011/12 looked after return (i.e. were looked after at some point over that school year), of whom 16,859 were of possible or definite school age (defined as aged 4 to 19 years inclusive at the start of the school year). Of the 16,859, 13,357 (79.2%) children had any SCN recorded on their looked after record and 11,182 (66.3%) had a valid SCN recorded. The completeness of SCN recording varied by local authority. Whilst most authorities returned a valid SCN on at least 80% of looked after records for children of school age a minority had noticeably poorer completeness, for example 18% in Dumfries & Galloway and 26% in Glasgow City.

Table 1 about here.

The number of children in a publicly funded school in 2011/12 identified as currently or previously looked after, and the proportions for whom a safe link to the CHI database could

be made, are shown in Table 2 (see Supplementary material for a definition of ‘safe links’).
 In total, 670,952 children were included in the 2011/12 pupil census. Of the 11,182 valid
 SCNs included in the 2011/12 looked after returns, 10,009 were also included in the 2011/12
 pupil census and these children (1.5% of all pupils) were classified as currently looked after.
 SCNs may be included in the 2011/12 looked after returns but not the 2011/12 pupil census
 if: the SCN recorded on the looked after return is incorrect; the child has left publicly funded
 school and their historical SCN has been included on their looked after return, or; they are
 home schooled or attending an independent school but have an SCN due to being registered
 for examinations with the Scottish Qualifications Authority. An additional 1,757 (0.3%)
 children in the 2011/12 pupil census were classified as previously looked after, that is their
 SCN had been included in a previous (2007/08 to 2010/11) looked after return, but not the
 2011/12 return. The remaining 659,186 (98.2%) children were classified as non-looked after.
 Compared to non-looked after children, children in school in 2011/12 classified as looked
 after were more likely to be of compulsory secondary school age (12-15 years) rather than
 primary school (4-11) or post-compulsory secondary school (16-19) age. Looked after
 children were also more likely to be male, much more likely to live in a deprived area, and
 more likely to be of white British ethnicity (Table 2).
 A safe linkage to the CHI database was achieved for 95.1% (626,732/659,186) of non-looked
 after children. Linkage rates were marginally lower for children classified as currently
 looked after (94.0%, 9,409/10,009) and similar for children classified as previously looked
 after (95.3%, 1,674/1,757). Amongst non-looked after children, linkage rates to CHI were
 slightly higher for secondary compared to primary school aged children and for those living
 in more affluent areas. Similar patterns (with steeper discrepancies) were seen for looked

after children. Linkage rates showed only moderate variation by local authority for looked after and non-looked after children.

Table 2 about here.

The proportion of currently looked after children that could be safely linked to the CHI database by the characteristics of looked after care received is shown in Table 3. Linkage rates were lower for children experiencing a higher number of placements and for those in residential accommodation compared to those being looked after at home. Considering the statutory basis for children's care, linkage rates were lower for children with an emergency court order and higher for those with a permanence order (a legal instrument enabling a permanent fostering arrangement or adoption) compared to those under compulsory supervision at home.

Table 3 about here.

Discussion

We used routinely collected looked after children and pupil census data to categorise children in publicly funded schools in Scotland in 2011/12 as currently, previously, or non-looked after. Recording of children's Scottish Candidate Number on looked after data is incomplete hence some under-ascertainment of looked after children within the general school population will occur. We linked all children in the pupil census to the NHS Scotland master patient index (Community Health Index database) and demonstrated that a safe link to the child's CHI number could be obtained for a very high proportion of looked after and non-looked after children (94% and 95% of children classified as currently looked after and non-looked after respectively).

Strengths and limitations

Categorising children in the pupil census as currently, previously, or non-looked after

Recording of the Scottish Candidate Number on looked after records is incomplete. We found that overall 79.2% of children of possible or definite school age had an SCN on their looked after record in 2011/12, with completeness varying markedly between local authorities. The recording of SCNs on looked after records by local authority staff is a manual administrative process. The variation in data quality between local authorities is therefore a reflection of differing operational practices that affect both data completeness and accuracy. We would not expect 100% completeness from any area as, even within the 4-19 year old age group, some children will not have started school, some will have left school, and some may be home schooled or attending independent mainstream or special schools. Overall, fewer than 1% of all school children in Scotland are home schooled¹⁴ and around 4% attend an independent school¹⁵. The proportion of looked after children receiving these types of education is unknown, although it is likely, for example, that a number of looked after children are in independent specials schools.

We found that 16% of the SCNs that were included in the 2011/12 looked after records were invalid. When looked after records are returned from local authorities to the Scottish Government, all submitted SCNs are checked to ensure they are the correct length (nine digits) and contain a correct check digit. In addition, a check is run to ensure that each SCN only occurs once in the dataset submitted by any individual local authority. However, although any records failing the validation checks are brought to the attention of the submitting local authority, if it is not possible to correct an error the record is still accepted into the national looked after dataset.

Incomplete or inaccurate recording of SCN on looked after returns means that some children within the 2011/12 pupil census who were looked after during that academic year will have been misclassified as non-looked after. It is not possible to precisely quantify the extent of this under-ascertainment of 'currently looked after' children. The published national statistics on looked after children in Scotland suggest that as at mid 2015, 1.6% (11,990 of 739,922) of the school aged population (children 5-17 years) were looked after⁶. We classified 1.5% of children in the 2011/12 pupil census as looked after at some point over that academic year, suggesting only a modest degree of under-ascertainment of our currently looked after group.

The degree of under-ascertainment of children in school in 2011/12 that were previously looked after is likely to be substantially higher than that for currently looked after children. Using all available looked after returns we estimate that approximately 13,000 school aged children had been looked after at some point over the academic years 2007/08 to 2010/11 (but not 2011/12) and were still school aged during 2011/12. This provides a maximum estimate of the number of previously looked after children we may have anticipated identifying in the 2011/12 pupil census. In practice however we only identified 1,757 children in the 2011/12 pupil census as previously looked after. This is likely to reflect a number of factors including that some of the 13,000 were not in a publicly funded school either at the time they were looked after (for example they had not yet started school) or during 2011/12 (for example they had left school early), or that their SCN was missing or inaccurately recorded on their historical looked after record hence could not be identified in the 2011/12 pupil census. The proportion of looked after returns for school aged children that contained any SCN was 67% in 2007/08 (the first year covered by the return) and 53% in

2008/09. Since 2009/10 it has been consistently higher than 75%, although further improvement has been lacking in recent years.

Overall, it is highly likely that children in school in 2011/12 that we identified as currently or previously looked after were correctly classified however some of the children identified as non-looked after will in fact have been currently or, more likely, previously looked after. Any such misclassified children are likely to comprise a very small proportion of all ‘non-looked after’ children. In addition, any such misclassification will tend to minimise any differences observed between looked after and non-looked after groups, i.e. will conservatively bias any findings of health differences between looked after and non-looked after groups towards the null.

Linking pupil census records to the Community Health Index

Pupil date of birth, gender, and home postcode are the only personal identifiers available within the pupil census. Individuals’ names, a key variable usually used when linking together health records belonging to the same people, are not available. In addition, looked after children are likely to have relatively high residential mobility, with frequent changes of home postcode. Despite these challenges, we found that linkage rates for children in the currently looked after group were only marginally lower than those achieved for non-looked after children, with a safe link to a Community Health Index number found for 94% and 95% respectively.

Linkage rates were generally highest for older school children, probably reflecting the availability of more years’ pupil census data for older children and hence potentially multiple postcodes that could increase the chance of a match to that held on the CHI database. Linkage rates were also generally higher for children living in less deprived areas, which is

likely to reflect lower residential mobility amongst affluent groups. Amongst currently looked after children, linkage rates showed some variation by placement number and type, again generally suggesting that more stable living arrangements were associated with higher linkage rates. The only exception to this was the relatively low linkage rate found for children living in state provided residential accommodation: the reason for this is unclear. Well established processes were used to link children's identifiers held on the pupil census to those in the CHI database. Of particular note, we have demonstrated previously that CHI numbers accepted as safe links for particular SCNs are highly likely (>99%) to be correct¹⁶.

The Scottish Candidate Number and Community Health Index number are both managed nationally. As children move between local authority or NHS Board areas they retain the same SCN and CHI number. Importantly when considering looked after children, children also generally retain the same SCN and CHI number following adoption.

Findings in light of existing literature

The potential contribution of population based studies involving the linkage of routine administrative data to elucidating the risk factors for, and outcomes from, state care in childhood has been well recognised^{17, 18}. Several examples of studies based on linkage of administrative state care/looked after children data are available. Studies linking to health records have examined: risk factors for entering state care such as maternal characteristics and neonatal health^{19, 20} and prior hospitalisation for injuries²¹; outcomes of children in care such as use of emergency/unscheduled health care²², self harm and suicide²³, and overall mortality²⁴; and secular trends in child maltreatment²⁵. Studies linking to data from other sectors have examined wider social outcomes for children in state care, in particular educational attainment^{26, 27}. Published linkage studies generally come from Australia, North

America, or Scandinavia and include a sub-national sample and/or lack non-looked after controls. We are not aware of any previous published linkage studies that provide whole population data on looked after and non-looked after children at national level.

Within the wider public health literature there is increasing recognition of the potential for incomplete or inaccurate data linkage to introduce considerable bias into studies²⁸⁻³⁰ and reporting guidelines for data linkage studies have been suggested^{31, 32}. To date, however, studies reporting linked analyses of looked after children data generally provide minimal information on the linkage process or the quality of linkage achieved, and no discussion of potential bias that may be introduced through differential linkage rates for different groups, focusing instead on reporting the results of linked analyses.

Conclusions and recommendations

We have shown that it is possible using currently available routine looked after and education data to categorise children in publicly funded schools in Scotland into currently, previously, and non-looked after groups. Some under-ascertainment of looked after groups occurs, with some (particularly previously) looked after children misclassified as non-looked after.

Children categorised to looked after and non-looked after groups can be linked to their Community Health Index numbers (and hence to all routinely collected health data) with very high completeness and accuracy. We demonstrate some variation in linkage quality by factors such as local authority area and the complexity of children's looked after journeys.

Overall our results suggest that linkage of currently available routine looked after, education, and health data is feasible and likely to provide linked data fit for epidemiological research purposes. We would caution that researchers working with linked data should always

scrutinise the quality of the underlying linkage and consider the extent to which bias introduced through the linkage process may exaggerate or minimise findings, particularly when making comparisons across subgroups. In general however, the health outcomes of looked after children are much poorer than those of non-looked after children hence differences are likely to be readily evident. This was demonstrated in our follow on analysis which showed much poorer dental health for looked after compared to non-looked after children, including after accounting for age, sex and deprivation differences.

Improvement to the completeness and accuracy of recording of the Scottish Candidate Number on looked after returns would improve the classification of children in the pupil census as currently, previously, and non-looked after. Due to the limited personal identifiers available on looked after records, it is not currently possible to link looked after data directly to the Community Health Index: the intermediate step of the pupil census is required. This means that it is currently not possible to link looked after records for preschool children, those who have left school, or those in non publicly funded schools to the children's health data. Expanding the identifiers included on looked after returns to allow direct linkage to the CHI database, or indeed inclusion of a unique 'citizen number' on routine administrative data from all sectors, would present further opportunities to explore the health and healthcare of these groups, although any such developments would need to carefully balance the benefits of enhanced analysis opportunities against potential privacy risk³³.

Overall this novel linkage presents opportunities to develop further research on the health and healthcare of school aged looked after children compared to their peers, enabling scrutiny of whether recent policy and practice developments are resulting in discernible improvements and reduced inequalities. Researchers wishing to access linked routine health and looked

408 after data should contact the NHS Scotland Information Services Division research support
409 service in the first instance³⁴.

410 **Declarations**

411 **Ethics approval**

412 Approval for this study was obtained from the Scottish Government Education Analytical
413 Services Division Data Access Panel and the NHS Privacy Advisory Committee (now the
414 Public Benefit and Privacy Panel). The approval processes involved completion of a privacy
415 impact assessment, a data sharing agreement between the Scottish Government and the
416 University of Glasgow, and a data processing agreement between the University of Glasgow
417 and the NHS Information Services Division. Ethical approval was obtained from the
418 University of Glasgow College of Medicine, Veterinary, and Life Sciences Ethics
419 Committee: NHS ethical approval was not required.

420 **Consent for publication**

421 Not applicable.

422 **Availability of data and materials**

423 The Pupil Census and Children Looked After Survey (CLAS) datasets are held by the
424 Education Analytical Services Directorate within the Scottish Government. Information on
425 how to request permission to access these datasets for research purposes is described here:
426 <http://www.gov.scot/Topics/Statistics/Browse/School-Education/DataAccess>.
427 An analysis copy of the Community Health Index (CHI) database is held by NHS National
428 Services Scotland Information Services Division (ISD). How to request permission to access
429 the stand alone CHI dataset (or linked health, education, and children's social work data) is
430 described here <http://www.isdscotland.org/Products-and-Services/eDRIS/>.

Funding

This study was funded by a National Records of Scotland Cross-Sectoral Data Linkage Pathfinder grant (2013-2016). The funding body had no role in the design, conduct, or reporting of the study.

Authors' contributions

DC provided CHI data and conducted the linkage and statistical tests. AK provided pupil census and looked after data and contributed to conducting the linkage. KS led on securing governance permissions. RW wrote the manuscript. All authors contributed to study design and reviewing the manuscript. All authors approved the final submission.

Competing interests

No authors have any competing interest to declare.

Acknowledgements

We acknowledge the contribution of Winifried van der Sluijs and Carrie Graham to the project steering group.

List of abbreviations

CHI: Community Health Index

CLAS: Children Looked After Survey

ISD: NHS National Services Scotland Information Services Division

LA: Local Authority

NHS: National Health Service

SCN: Scottish Candidate Number

SIMD: Scottish Index of Multiple Deprivation

References

1. UK Parliament. *Children (Scotland) Act 1995*. London: HMSO; 1995.
2. Scottish Parliament. *The Looked After Children (Scotland) Regulations 2009*. Edinburgh: The Stationery Office; 2009.
3. Scottish Parliament. *Children and Young People (Scotland) Act 2014*. Edinburgh: The Stationery Office; 2014.
4. About Children's Hearings Scotland. <http://www.chscotland.gov.uk/about-chs/>. Accessed 28 November 2016.
5. Audit Scotland. *Getting it right for children in residential care*. Edinburgh: Audit Scotland; 2010.
6. Scottish Government. *Children's social work statistics Scotland, 2014-15*. Edinburgh: Scottish Government; 2016.
7. Scottish Executive. *Looked after children and young people: we can and must do better*. Edinburgh: Scottish Executive; 2007.
8. Scott J, Hill M. *The health of looked after and accommodated children and young people in Scotland: messages from research*. Edinburgh: Social Work Inspection Agency; 2006.
9. Scott S, Hattie R, Tannahill C. *Looked after children in Glasgow and Scotland: a health needs assessment*. Glasgow: Scottish Public Health Network; 2013.
10. Scottish Government. *Getting it right for looked after children and young people: early engagement, early permanence, and improving the quality of care*. Edinburgh: Scottish Government; 2015.
11. Scottish Government. *Looked after children data strategy, 2015*. Edinburgh: Scottish Government; 2015.
12. Scottish Government. *Education outcomes for looked after children 2014/15*. Edinburgh: Scottish Government; 2016.
13. The Scottish Index of Multiple Deprivation. <http://www.gov.scot/Topics/Statistics/SIMD>. Accessed 28 November 2016.
14. Schoolhouse. <http://www.schoolhouse.org.uk/>. Accessed 28 November 2016.
15. Scottish Government. *Independent school census, September 2009*. Edinburgh: Scottish Government; 2010.
16. Wood R, Clark D, King A, Mackay D, Pell J. Novel cross-sectoral linkage of routine health and education data at an all-Scotland level: a feasibility study. *Lancet* 2013;382(S10):10.

- 486 17. Brownell MD, Jutte DP. Administrative data linkage as a tool for child maltreatment
487 research. *Child Abuse Neglect* 2013;**37**:120-4.
- 488 18. Putnam-Horstein E, Needell B, Rhodes AE. Understanding risk and protective factors for
489 child maltreatment: the value of integrated, population based data. *Child Abuse Neglect*
490 2013;**37**:116-9.
- 491 19. Kalland M, Sinkkonen J, Gissler M, Merilainen J, Siimes MA. Maternal smoking
492 behaviour, background and neonatal health in Finnish children subsequently placed in
493 foster care. *Child Abuse Neglect* 2006;**30**:1037-47.
- 494 20. Needell B, Barth RP. Infants entering foster care compared to other infants using birth
495 status indicators. *Child Abuse Neglect* 1998;**22**:1179-87.
- 496 21. O'Donnell M, Nassar N, Leonard H, Jacoby P, Mathews R, Patterson Y, Stanley F. Rates
497 and types of hospitalisations for children who have subsequent contact with the child
498 protection system: a population based case-control study. *J Epidemiol Commun H*
499 2010;**64**:784-8.
- 500 22. Rubin DM, Alessandrini EA, Feudtner C, Localio AR, Hadley T. Placement changes and
501 emergency department visits in the first year of foster care. *Pediatrics* 2004;**114**:e354-60.
- 502 23. Katz LY, Au W, Singal D, Brownell M, Roos N, Martens P, Chateau D, Enns MW,
503 Kozyrskyj A, Sareen J. Suicide and suicide attempts in children and adolescents in the
504 child welfare system. *Can Med Assoc J* 2011;**183**:1977-81.
- 505 24. Jonson-Reid M, Chance T, Drake B. Risk of death among children reported for nonfatal
506 maltreatment. *Child Maltreatment* 2007;**12**:86-95.
- 507 25. Gilbert R, Fluke J, O'Donnell M, Gonzalez-Izquierdo A, Brownell M, Gulliver P, Janson
508 S, Sidebotham P. Child maltreatment: variations in trends and policies in six developed
509 countries. *Lancet* 2012;**379**:758-72.
- 510 26. Brownell M, Roos N, MacWilliam L, Leclair L, Ekuma O, Fransoo R. Academic and
511 social outcomes for high risk youths in Manitoba. *Can J Education* 2010;**33**:804-36.
- 512 27. Berger LM, Cancian M, Han E, Noyes J, Rios-Salas V. Children's academic achievement
513 and foster care. *Pediatrics* 2015;**135**:e109-16.
- 514 28. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, Brand CA. Data
515 linkage: a powerful research tool with potential problems. *BMC Health Serv Res*
516 2010;**10**:346.
- 517 29. Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of
518 record linkage. *J Epidemiol Commun H* 2012;**66**:1198.
- 519 30. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, Parslow R:
520 Data linkage errors in hospital administrative data when applying a pseudonymisation
521 algorithm to paediatric intensive care records. *BMJ Open* 2015;**5**:e008118.

- 522 31. Bohensky MA, Jolley D, Sundararajan V, Evans S, Ibrahim J, Brand C. Development
523 and validation of reporting guidelines for studies involving data linkage. *Aust NZ J Publ*
524 *Heal* 2011;**35**:486-9.
- 525 32. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sorensen HT,
526 von Elm E, Langan SM. The REporting of studies Conducted using Observational
527 Routinely collected health Data (RECORD) statement. *PLOS Med* 2015;**12**:e1001885.
- 528 33. Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data - a best practice
529 protocol. *Aust NZ J Publ Heal* 2002;**26**:251-5.
- 530 34. ISD: Electronic data research and innovation service.
531 <http://www.isdscotland.org/Products-and-Services/eDRIS/>. Accessed 28 November
532 2016.
- 533 35. ScotXed: Looked after children.
534 [http://www.gov.scot/Topics/Statistics/ScotXed/ChildrenandYoungPeople/LookedAfterC](http://www.gov.scot/Topics/Statistics/ScotXed/ChildrenandYoungPeople/LookedAfterChildren)
535 [hildren](http://www.gov.scot/Topics/Statistics/ScotXed/ChildrenandYoungPeople/LookedAfterChildren). Accessed 28 November 2016.
- 536 36. ScotXed: School/pupil census.
537 <http://www.gov.scot/Topics/Statistics/ScotXed/SchoolEducation/SchoolPupilCensus>.
538 Accessed 28 November 2016.
- 539 37. Scottish Government eHealth Division. *The use of the CHI (Community Health Index) to*
540 *support integrated care across the NHS in Scotland*. Edinburgh: Scottish Government;
541 2013.
- 542 38. NHS Scotland. *CHI User Manual. Appendix A: Data Contents. Version 9.8*. Edinburgh:
543 NHS National Services Scotland; 2012.
- 544 39. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. *Int J*
545 *Epidemiol* 2016;**45**:954-64.
- 546 40. Kendrick S, Clarke J. The Scottish record linkage system. *Health Bulletin* 1993;**51**:72-9.
- 547 41. Kendrick S. The development of record linkage in Scotland: the responsive application
548 of probability matching. In: *Federal Committee on Statistical Methodology. Record*
549 *Linkage Techniques - 1997*. Proceedings of an International Workshop and Exposition
550 March 20-21st 1997. Arlington, VA. Washington DC: Office of Management and the
551 Budget, US Government; 1997. p. 319-32.
- 552 42. Mackay DF, Wood R, King A, Clark D, Cooper SA, Smith GCS, Pell JP. Educational
553 outcomes following breech delivery: a record-linkage study of 456,947 children. *Int J*
554 *Epidemiol* 2015;**44**(1):209-17.

555

556 **Panel**

557 **Panel 1: Databases used in the linkage**

558 **Figures**

559 **Figure 1: Description of linkage process**

560

561 Legend for Figure 1

562 SCN – Scottish Candidate Number

563 CHI – Community Health Index

564

565 **Tables**

Table 1: Quality of recording of Scottish Candidate Number on looked after children data

	Looked after returns for 2011/12 for school aged children*				
	Total	Containing any SCN		Containing a valid SCN	
	Number (%)	Number	%	Number	%
All school aged looked after children					
Total	16,859 (100%)	13,357	79.2	11,182	66.3
Age group at start of school year					
4-11	8,592 (51%)	6,510	75.8	5,462	63.6
12-15	6,119 (36%)	5,149	84.1	4,320	70.6
16-19	2,148 (13%)	1,698	79.1	1,400	65.2
Gender					
Male	9,128 (54%)	7,224	79.1	6,066	66.5
Female	7,731 (46%)	6,133	79.3	5,116	66.2
Local authority area responsible for child					
Aberdeen City	641 (4%)	626	97.7	567	88.5
Aberdeenshire	541 (3%)	496	91.7	455	84.1
Angus	291 (2%)	283	97.3	258	88.7
Argyll & Bute	241 (1%)	227	94.2	208	86.3
Clackmannanshire	206 (1%)	193	93.7	177	85.9
Dumfries & Galloway	456 (3%)	94	20.6	83	18.2
Dundee City	723 (4%)	660	91.3	602	83.3
East Ayrshire	495 (3%)	408	82.4	352	71.1
East Dunbartonshire	164 (1%)	160	97.6	134	81.7
East Lothian	217 (1%)	214	98.6	194	89.4
East Renfrewshire	192 (1%)	146	76.0	138	71.9
Edinburgh, City of	1,381 (8%)	1,235	89.4	1,104	79.9
Falkirk	458 (3%)	421	91.9	377	82.3
Fife	842 (5%)	634	75.3	557	66.2
Glasgow City	3,694 (22%)	1,961	53.1	950	25.7
Highland	563 (3%)	496	88.1	451	80.1
Inverclyde	349 (2%)	314	90.0	281	80.5
Island Councils	124 (1%)	119	96.0	108	87.1
Midlothian	294 (2%)	280	95.2	244	83.0
Moray	227 (1%)	178	78.4	154	67.8
North Ayrshire	575 (3%)	427	74.3	386	67.1
North Lanarkshire	810 (5%)	756	93.3	684	84.4
Perth & Kinross	232 (1%)	219	94.4	201	86.6

Renfrewshire	784 (5%)	755	96.3	673	85.8
Scottish Borders	212 (1%)	198	93.4	179	84.4
South Ayrshire	371 (2%)	364	98.1	324	87.3
South Lanarkshire	643 (4%)	395	61.4	358	55.7
Stirling	274 (2%)	261	95.3	234	85.4
West Dunbartonshire	416 (2%)	411	98.8	373	89.7
West Lothian	443 (3%)	426	96.2	376	84.9

*School aged children defined as 5-20 years inclusive at the end of the period covered by the looked after return (July 2012) hence assumed to be 4-19 years inclusive at the start of the school year covered (August 2011 to July 2012)

Island Councils comprise Orkney Islands, Shetland Islands, and Na h-Eileanan Siar

571 **Table 2: Linkage rate to CHI database for currently, previously, and non-looked after children in 2011/12 pupil census**

	Currently looked after				Previously looked after				Non-looked after			All children		
	In pupil census	Linked to CHI			In pupil census	Linked to CHI			In pupil census	Linked to CHI		In pupil census	Linked to CHI	
	N (%)	N	%	Difference (95% CI)*	N (%)	N	%	Difference (95% CI)*	N (%)	N	%	N (%)	N	%
All children in pupil census														
Total	10,009 (100%)	9,409	94.0	-1.1 (-1.5, -0.6)	1,757 (100%)	1,674	95.3	0.2 (-0.8, 1.2)	659,186 (100%)	626,732	95.1	670,952 (100%)	637,815	95.1
Age group at start of school year														
4-11	4,399 (44%)	4,121	93.7	-0.8 (-1.5, -0.1)	924 (53%)	868	93.9	-0.5 (-2.1, 1.0)	384,674 (58%)	363,475	94.5	389,997 (58%)	368,464	94.5
12-15	5,235 (52%)	4,930	94.2	-1.7 (-2.4, -1.1)	731 (42%)	708	96.9	0.9 (-0.3, 2.2)	216,430 (33%)	207,596	95.9	222,396 (33%)	213,234	95.9
16-19	375 (4%)	358	95.5	-0.4 (-2.6, 1.7)	102 (6%)	98	96.1	0.2 (-3.6, 3.9)	58,027 (9%)	55,651	95.9	58,504 (9%)	56,107	95.9
Gender														
Male	5,319 (53%)	5,018	94.3	-0.7 (-1.3, -0.1)	931 (53%)	886	95.2	0.1 (-1.2, 1.5)	335,665 (51%)	319,004	95.0	341,915 (51%)	324,908	95.0
Female	4,690 (47%)	4,391	93.6	-1.5 (-2.2, -0.8)	826 (47%)	788	95.4	0.3 (-1.1, 1.7)	323,521 (49%)	307,728	95.1	329,037 (49%)	312,907	95.1
SIMD deprivation quintile														
1 (most deprived)	4,224 (42%)	3,947	93.4	-1.2 (-1.9, -0.4)	684 (39%)	641	93.7	-0.9 (-2.7, 0.9)	137,865 (21%)	130,450	94.6	142,773 (21%)	135,038	94.6
2	2,247 (22%)	2,131	94.8	-0.3 (-1.3, 0.6)	514 (29%)	499	97.1	1.9 (0.5, 3.4)	125,022 (19%)	118,979	95.2	127,783 (19%)	121,609	95.2
3	1,606 (16%)	1,517	94.5	-0.8 (-1.9, 0.3)	297 (17%)	283	95.3	0.0 (-2.4, 2.4)	129,743 (20%)	123,592	95.3	131,646 (20%)	125,392	95.2
4	1,225 (12%)	1,152	94.0	-1.3 (-2.6, 0.0)	159 (9%)	152	95.6	0.3 (-2.9, 3.5)	136,508 (21%)	130,144	95.3	137,892 (21%)	131,448	95.3
5 (least deprived)	620 (6%)	596	96.1	0.3 (-1.2, 1.8)	93 (5%)	89	95.7	-0.1 (-4.2, 4.0)	127,032 (19%)	121,713	95.8	127,745 (19%)	122,398	95.8

Ethnicity														
White British	9,422 (94%)	8,861	94.0	-1.3 (-1.8, -0.8)	1,676 (95%)	1,595	95.2	-0.2 (-1.2, 0.9)	592,412 (90%)	564,718	95.3	603,510 (90%)	575,174	95.3
Other	587 (6%)	548	93.4	0.5 (-1.5, 2.5)	81 (5%)	79	97.5	4.7 (1.3, 8.0)	66,774 (10%)	62,014	92.9	67,442 (10%)	62,641	92.9
Local authority area of school														
Aberdeen City	372 (4%)	355	95.4	0.0 (-2.1, 2.2)	84 (5%)	82	97.6	2.2 (-1.0, 5.5)	20,975 (3%)	20,009	95.4	21,431 (3%)	20,446	95.4
Aberdeenshire	427 (4%)	400	93.7	-2.1 (-4.4, 0.2)	75 (4%)	72	96.0	0.2 (-4.2, 4.7)	33,508 (5%)	32,095	95.8	34,010 (5%)	32,567	95.8
Angus	212 (2%)	205	96.7	0.5 (-1.9, 2.9)	28 (2%)	28	100.0	3.8 (3.5, 4.1)	14,998 (2%)	14,427	96.2	15,238 (2%)	14,660	96.2
Argyll & Bute	183 (2%)	168	91.8	-2.3 (-6.3, 1.7)	18 (1%)	18	100.0	5.9 (5.4, 6.3)	10,878 (2%)	10,238	94.1	11,079 (2%)	10,424	94.1
Clackmannanshire	154 (2%)	146	94.8	-0.6 (-4.1, 3.0)	19 (1%)	16	84.2	-11.2 (-27.6, 5.2)	6,461 (1%)	6,163	95.4	6,634 (1%)	6,325	95.3
Dumfries & Galloway	164 (2%)	156	95.1	-1.2 (-4.5, 2.1)	108 (6%)	97	89.8	-6.5 (-12.2, -0.8)	19,087 (3%)	18,385	96.3	19,359 (3%)	18,638	96.3
Dundee City	415 (4%)	397	95.7	-0.6 (-2.6, 1.4)	141 (8%)	134	95.0	-1.2 (-4.8, 2.4)	16,759 (3%)	16,133	96.3	17,315 (3%)	16,664	96.2
East Ayrshire	314 (3%)	289	92.0	-3.5 (-6.5, -0.5)	44 (3%)	43	97.7	2.2 (-2.2, 6.6)	15,768 (2%)	15,060	95.5	16,126 (2%)	15,392	95.4
East Dunbartonshire	94 (1%)	84	89.4	-5.9 (-12.1, 0.4)	20 (1%)	20	100.0	4.8 (4.4, 5.1)	15,758 (2%)	15,005	95.2	15,872 (2%)	15,109	95.2
East Lothian	148 (1%)	141	95.3	-0.9 (-4.3, 2.6)	39 (2%)	38	97.4	1.3 (-3.7, 6.3)	13,173 (2%)	12,662	96.1	13,360 (2%)	12,841	96.1
East Renfrewshire	102 (1%)	100	98.0	3.6 (0.9, 6.3)	16 (1%)	15	93.8	-0.7 (-12.6, 11.2)	16,075 (2%)	15,182	94.4	16,193 (2%)	15,297	94.5

Edinburgh, City of	795 (8%)	750	94.3	-1.2 (-2.8, 0.4)	131 (7%)	127	96.9	1.4 (-1.5, 4.4)	43,533 (7%)	41,578	95.5	44,459 (7%)	42,455	95.5
Falkirk	364 (4%)	346	95.1	-1.2 (-3.4, 1.0)	66 (4%)	63	95.5	-0.8 (-5.8, 4.2)	20,308 (3%)	19,548	96.3	20,738 (3%)	19,957	96.2
Fife	621 (6%)	580	93.4	-2.2 (-4.1, -0.2)	222 (13%)	215	96.8	1.3 (-1.0, 3.6)	46,983 (7%)	44,906	95.6	47,826 (7%)	45,701	95.6
Glasgow City	1,479 (15%)	1,364	92.2	-0.8 (-2.2, 0.6)	38 (2%)	37	97.4	4.4 (-0.7, 9.4)	64,284 (10%)	59,794	93.0	65,801 (10%)	61,195	93.0
Highland	365 (4%)	343	94.0	-0.9 (-3.4, 1.5)	100 (6%)	93	93.0	-1.9 (-6.9, 3.1)	30,640 (5%)	29,072	94.9	31,105 (5%)	29,508	94.9
Inverclyde	235 (2%)	220	93.6	-0.3 (-3.4, 2.9)	34 (2%)	30	88.2	-5.7 (-16.5, 5.2)	10,060 (2%)	9,446	93.9	10,329 (2%)	9,696	93.9
Island Councils	84 (1%)	83	98.8	4.1 (1.7, 6.5)	12 (1%)	11	91.7	-3.1 (-18.7, 12.6)	9,430 (1%)	8,932	94.7	9,526 (1%)	9,026	94.8
Midlothian	210 (2%)	200	95.2	-0.2 (-3.1, 2.7)	67 (4%)	65	97.0	1.6 (-2.5, 5.7)	11,387 (2%)	10,864	95.4	11,664 (2%)	11,129	95.4
Moray	145 (1%)	139	95.9	0.3 (-3.0, 3.5)	22 (1%)	20	90.9	-4.7 (-16.7, 7.3)	11,759 (2%)	11,240	95.6	11,926 (2%)	11,399	95.6
North Ayrshire	321 (3%)	300	93.5	-1.8 (-4.5, 0.9)	64 (4%)	59	92.2	-3.0 (-9.6, 3.5)	18,182 (3%)	17,315	95.2	18,567 (3%)	17,674	95.2
North Lanarkshire	599 (6%)	565	94.3	-0.6 (-2.4, 1.3)	59 (3%)	54	91.5	-3.4 (-10.5, 3.7)	48,002 (7%)	45,550	94.9	48,660 (7%)	46,169	94.9
Perth & Kinross	168 (2%)	155	92.3	-3.2 (-7.3, 0.8)	19 (1%)	19	100.0	4.5 (4.2, 4.8)	17,258 (3%)	16,479	95.5	17,445 (3%)	16,653	95.5
Renfrewshire	469 (5%)	442	94.2	-0.9 (-3.0, 1.3)	52 (3%)	51	98.1	3.0 (-0.8, 6.7)	22,957 (3%)	21,832	95.1	23,478 (3%)	22,325	95.1
Scottish Borders	154	142	92.2	-3.3	17	15	88.2	-7.3	14,748	14,088	95.5	14,919	14,245	95.5

	(2%)			(-7.6, 0.9)	(1%)			(-22.6, 8.0)	(2%)			(2%)		
South Ayrshire	243 (2%)	232	95.5	0.2 (-2.4, 2.9)	37 (2%)	36	97.3	2.0 (-3.2, 7.3)	14,024 (2%)	13,359	95.3	14,304 (2%)	13,627	95.3
South Lanarkshire	409 (4%)	380	92.9	-1.7 (-4.2, 0.8)	71 (4%)	70	98.6	4.0 (1.2, 6.7)	42,585 (6%)	40,291	94.6	43,065 (6%)	40,741	94.6
Stirling	188 (2%)	178	94.7	-0.7 (-3.9, 2.5)	31 (2%)	31	100.0	4.6 (4.3, 5.0)	12,200 (2%)	11,634	95.4	12,419 (2%)	11,843	95.4
West Dunbartonshire	228 (2%)	218	95.6	1.4 (-1.3, 4.1)	57 (3%)	56	98.2	4.0 (0.6, 7.5)	12,008 (2%)	11,314	94.2	12,293 (2%)	11,588	94.3
West Lothian	347 (3%)	331	95.4	0.4 (-1.8, 2.6)	66 (4%)	59	89.4	-5.6 (-13.1, 1.8)	25,398 (4%)	24,131	95.0	25,811 (4%)	24,521	95.0

* Difference between linkage rate for currently/previously looked after children and that for non-looked after children with 95% confidence interval
SIMD is Scottish Index of Multiple Deprivation

Island Councils comprise Orkney Islands, Shetland Islands, and Na h-Eileanan Siar

572
573
574
575
576

577
578

Table 3: Linkage rate to CHI database for currently looked after children by characteristics of looked after care

	Currently looked after			
	In pupil census	Linked to CHI		
	N (%)	N	%	Difference (95% CI)*
All currently looked after children in pupil census				
Total	10,009 (100%)	9,409	94.0	n/a
Number of looked after placements since August 2007				
1	6,049 (60%)	5,715	94.5	reference
2	1,762 (18%)	1,654	93.9	-0.6 (-1.9, 0.7)
3	1,052 (11%)	978	93.0	-1.5 (-3.2, 0.1)
4+	1,146 (11%)	1,062	92.7	-1.8 (-3.4, -0.2)
Type of most recent placement				
Looked After at home	4,285 (43%)	4,042	94.3	reference
Kinship care	2,196 (22%)	2,064	94.0	-0.3 (-1.6, 0.9)
Foster carers/other community	2,650 (26%)	2,507	94.6	0.3 (-0.8, 1.4)
Residential accommodation	878 (9%)	796	90.7	-3.7 (-5.7, -1.6)
Legal reason underpinning most recent looked after episode				
Compulsory supervision through the Children's Hearing system (resident at home)	3,961 (40%)	3,732	94.2	reference
Accommodated away from home on a voluntary basis	914 (9%)	850	93.0	-1.2 (-3.0, 0.6)
Compulsory supervision through the Children's Hearing system (resident away from home)	3,415 (34%)	3,216	94.2	0.0 (-1.1, 1.0)
Emergency court order	189 (2%)	169	89.4	-4.8 (-9.2, -0.4)
Permanence order or awaiting adoption	675 (7%)	650	96.3	2.1 (0.5, 3.7)
Other/unknown	855 (9%)	792	92.6	-1.6 (-3.5, 0.3)

* Difference between linkage rate for particular category and the relevant reference category with 95% confidence interval

579
580

Children Looked After Survey (CLAS) annual return: 'looked after data'

Since 2007/08, all 32 Scottish local authorities have been required to submit individual level data on looked after children to the Scottish Government³⁵. An annual return is made in November each year, providing information on all children looked after by the local authority at any point over the preceding school year (August to July).

Limited personal identifiers are included in the looked after returns, specifically the child's date of birth, gender, ethnicity, their social work number (specific to that local authority only), and, if available, their Scottish Candidate Number (see below). Children's names and home postcodes are not included.

Basic information on all episodes of looked after care provided by the local authority to the child (from birth) is also included, for example the start and end dates, legal reason, and type (e.g. at home, kinship care, foster care, etc.) of each placement.

Pupil census annual return: 'pupil census'

Since 2006/07, all 32 Scottish local authorities have been required to submit individual level data on children attending publicly funded schools to the Scottish Government³⁶. An annual return is made in October each year, providing information on children in school at the start of that school year (census date mid September). Publicly funded schools comprise mainstream and special schools funded by the state and managed by local authorities. Children who are home schooled or attending an independent or charitable day or residential school or secure unit will not be included in the pupil census, even if their school place is being funded by their local authority. A wider range of personal identifiers is included in the pupil census return, specifically the child's date of birth, gender, home postcode, ethnicity, and their Scottish Candidate Number (SCN). Pupil names are not included. The SCN is the unique pupil identifier used across Scotland. Pupils are assigned an SCN when they join a publicly funded school (typically at age 4/5 years) and/or are registered to undertake assessments administered by the Scottish Qualifications Authority. The SCN is recorded on all national level education datasets held by the Scottish Government, for example those on pupil attendance, exclusions, attainment, and post school destination.

Basic information on the education of each child is included in the pupil census, for example the school and school year attended and any recognised additional educational support needs.

Community Health Index (CHI) database: 'CHI database'/'health data'

The CHI database is NHS Scotland's master patient index^{37, 38}. It was introduced in one area of Scotland in the 1970s and has had national coverage since 1997. It covers all patients registered with a General Practitioner and others provided with care from NHS Scotland, for example individuals attending emergency services.

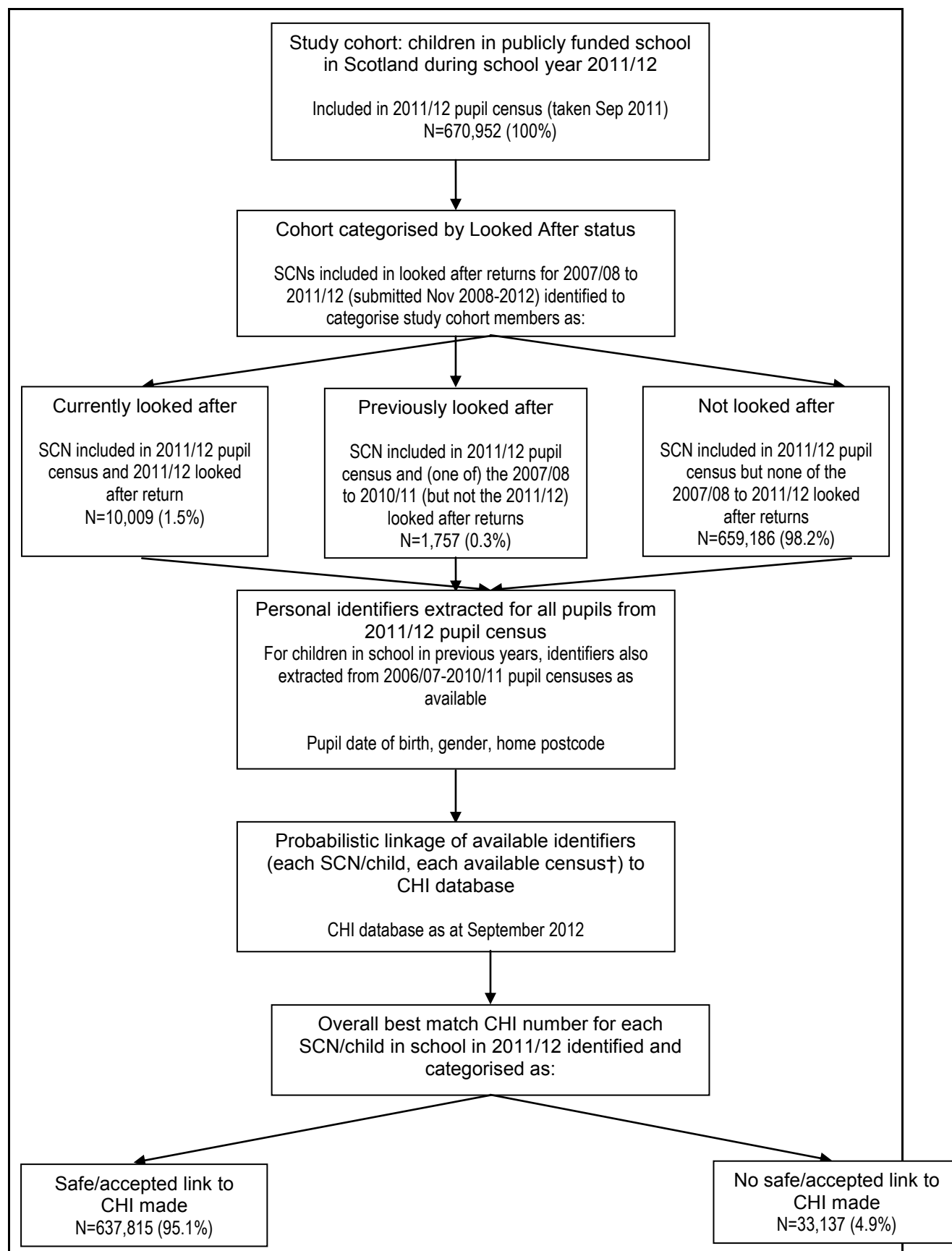
Full personal identifiers are included in the database, specifically patient full name, date of birth, gender, home postcode and their CHI number. The database is 'live' and updated as required by authorised staff. The NHS Information Services Division (ISD) receives a monthly snapshot download of the database for statistical and data linkage purposes. The CHI number is the unique patient identifier used in NHS Scotland. The CHI number is held on all national level health datasets held by ISD, for example those on hospital admission, Accident & Emergency department and outpatient attendance, community prescriptions, and child health reviews and vaccinations.

Very limited information on health care is included in the CHI database, for example the patient's registered General Practice.

Data quality

For each of the databases, validation is built into the data recording/submission process.

For the looked after and pupil census returns, additional data quality assurance is undertaken by the Scottish Government and local authorities are required to confirm the validity of returns.



Supplementary material

We used standard probabilistic record linkage methods to link pupil census records to the CHI database³⁹. The algorithms and decision rules used are used by NHS Information Services Division (ISD) for routine linkage of health records^{40, 41} and have previously been extended to link education and health records^{16, 42}. The algorithms return a suggested overall 'best match' CHI number for each SCN/child within the pupil census and assign each best match CHI number to a predetermined match category based on: (a) the closeness of match between each of the personal identifiers held on the pupil census and the CHI database for the respective SCN and CHI numbers and, (b) how closely the 'next best match' CHI number suggested for a particular SCN rivals the 'best match'. Best match CHI numbers assigned to certain pre-specified match categories were accepted as safe links for the relevant SCNs. Best match CHI numbers assigned to other match categories were rejected and the relevant SCNs were considered not to link to a CHI number (see Supplementary table 1).

As shown in Figure 1, we had personal identifiers (date of birth, sex, and home postcode) for all children in the 2011/12 pupil census and this group formed our study population. In addition, we also received pupil census records for each of the previous five years (2006/07 to 2010/11) each containing the personal identifiers for each child/SCN at the time of that census. Therefore, pupils who were older could have up to six sets of identifiers (including, for example, up to six home postcodes) associated with their SCN for linkage to the CHI. Different postcodes included in the census for a particular child in different years could reflect genuine change (i.e. the child had moved house during the year) or data error (i.e. one or other of the recorded postcodes was erroneous).

711 The CHI database snapshot, taken for linkage to the pupil census records in September 2012,
712 contained for every person in the database at least their current home postcode as recorded at
713 their GP Practice as well as their last previous known postcode. In some instances, for
714 example where a patient had moved NHS Board boundaries within Scotland, additional CHI
715 records for that person would capture further postcodes, which were also available for
716 linkage.

717

718 Each set of pupil census identifiers for each SCN/child was linked to the CHI database as at
719 Sep 2012 to generate up to six 'best match' CHI numbers for each child. The postcode
720 recorded within any census record could match to any of the postcodes available within a
721 CHI record. The overall best match CHI for each SCN/child was then taken as that with the
722 highest overall match category. In case of ties, the best match CHI from the most recent
723 census year was taken as the overall best match for that child.

724

725 Making use of all available identifiers, in particular multiple postcodes, in this way provides
726 an opportunity to overcome administrative mismatches between the pupil census and CHI
727 database. For example, if a GP practice was unaware that a child had moved house, a match
728 may be made between the postcode held on a previous pupil census to that on the CHI
729 database. This flexibility may be particularly important for looked after children who have
730 high residential mobility and multiple possible 'home' addresses.

731

732 Since the linkage for this study was carried out, ISD has made additional enhancements to the
733 processes used when linking education and health records. For example, ISD can now
734 capture patients' full postcode histories from the CHI database (see

735 <http://www.isdscotland.org/Products-and-Services/eDRIS/Docs/20150421-Linking-ScotXed->

736 [Data.pdf](#)). Further developments are underway involving linkage to the national indexing
737 spine and establishment of 'read-through' indexes to allow more efficient regular linkage of
738 education and health records.

739

Supplementary table 1: Linkage of pupil census records to CHI database: detailed results

Match category	Best match CHI for a particular SCN: match between personal identifiers held on pupil census and CHI database			Best match CHI compared to next best match CHI for a particular SCN	Best match CHI considered a safe/acceptable link for the SCN?	Number of children in 2011/12 pupil census in this category
	Date of birth	Gender	Postcode			
A	Exact	Exact	Exact	Unrivalled	✓	87,259
B	Exact	Exact	Exact	Distant rival	✓	482,325
C	Exact	Exact	Exact	Intermediate rival	✓	53,024
D	Exact	Exact	Exact	Close rival	✓	6,153
E	Exact	Exact	Exact	Tied	✗	12,959
F	Exact	Exact	Close	Unrivalled	✓	1,167
G	Exact	Exact	Close	Rivalled	✓	7,887
H	Exact	Exact	Close	Tied	✗	239
I	Close	Exact	Exact	Unrivalled	✗	332
J	Close	Exact	Exact	Rivalled	✗	1,428
K	Close	Exact	Exact	Tied	✗	26
L	Other combination of close matches/high overall linkage score			Unrivalled	✗	1,239
M	Other – considered a non-match				✗	16,914

Close match on date of birth indicates 2 out of 3 (of DD, MM, YY) agreed; close match on postcode indicates 6 out of 7 characters agreed

Unrivalled means the next best CHI had a much lower linkage score than the best match CHI

Tied means that the next best CHI had the same linkage score as the best match CHI; other categories are intermediate

740

741

742

743

744